Cognitive evolution of transport spatiality

Jonas Matijošius
Vilnius Gediminas Technical University
Vilnius, Lithuania
jonas.matijosius@vliustech.lt

Abstract
This study aims to provide an overview of the cognitive evolution of transport related to spatial theorems. The study recalls the various economic theories and criticizes them. It also reviews their methodology and their transport connection, with spatial focus. It can be stated that in the last two hundred years, spatiality appeared in transport related economics modelling, meanwhile sustainability or environmental protection are not in the scope of such models yet. Therefore, new area is to emerge, in which sustainability will be in focus.

Keywords
Transport economics, Spatial economics, Cognitive development, Theorem evolution, Sustainability

1. Simple models– the one-point economy

In this paper, the cognitive evaluation of transport spatial models was analysed, so the theoretical development, the new and innovative ideas and their criticism are presented. Principles of Economics is a leading economics textbook of Alfred Marshall (1842–1924), originally a mathematician, which was first published in 1890 (Marshall, 1890). According to Marshall’s theory, industries develop in areas with a priori good endowments (e.g., raw materials), and then clusters are formed due to agglomeration advantages (local specialisation, presence of service industries). He also described internal and external economies of scale. In a one-point economy, it is assumed that supply and demand meet, so that unit prices and average costs are established. The spatial location of economic agents is not meaningful. This model is based on the single space economy, no spatiality was considered, and it does not take into account competition. So in this model, the market is perfect, consumer behaviour is rational, and consumer preferences are homogeneous. The model was modified and improved several times after Ricardo, but its synthesis had to wait until the 1900s, when Heckscher and Ohlin modified it (1933).

David Ricardo published his paper on comparative advantage in 1817 (Ricardo, 2020). For international trade to be beneficial to all countries involved, the basic condition for a beneficial division of labour is that the productivity and efficiency rates of assorted products and sectors differ across countries. The condition for trade is therefore not that the endowments or productivity in one country are higher in absolute terms than in another, but that the relative efficiency of a given sector relative to other sectors in one country is more favourable than the relative position of the same sector in another country. If each country specialises in activities that are more efficient than its own average economic activity, division of labour will lead to greater overall efficiency and surplus production. This model is based on labour theory and does not consider other factors of production, and thus fails to reveal the real drivers and laws of the territorial division of labour. The model was modified and improved several times, but its synthesis had to wait until the 1900s (Heckscher, Ohlin).

In his 1826 work The Isolated City [in German: Der isolierte Staat], Johann Heinrich von Thünen (von Thünen, 1826) described his model of an idealised city, which later became the basis for other theories of spatial economics – a model that focuses on agriculture. Von Thünen makes the following assumptions in his model.

1. The city is completely isolated and surrounded by wilderness (it can only be cultivated at a loss because of its remoteness).
2. The soil quality and climate are uniform.
3. The price of products is the same everywhere.
4. There are no roads (transport by ox cart), the land is completely flat, there are no mountains or rivers.
5. Farmers behave rationally (profit maximisers), transport products to market by ox cart (transport costs increase linearly with distance).
Under these conditions, it is clear that farmers’ profits decrease with distance from the market and that the efficiency of production depends solely on this distance. Accordingly, it makes sense to produce products with a high volume and weight close to the market and to use areas further away to produce products with higher transport costs.

Revenues decrease with distance from the market and eventually equal the sum of transport and production costs – beyond this point production would be loss-making. Several types of goods can be produced in one area (city). In the idealized model, the most profitable product is produced at any distance from the market, and the boundary between each product will be where their profitability is equal, thus forming circles around the city centre. In von Thünen’s model, the circles follow each other in the following order:

![Diagram showing the isolation of city with concentric rings for different products](https://doi.org/10.55343/CogSust.33

![Figure 1. Visualisation of the isolated city (based on: von Thünen, 1826)](https://doi.org/10.55343/CogSust.33

1. city centre;
2. dairy products, vegetables (short distance only);
3. forestry (heavy weight, important energy source);
4. cereals;
5. livestock farming (“self-supplied” to the market, low transport costs).

In the model, the value of the land (“locational rent”) is the amount that the farmer can pay without losses. Von Thünen later extended his model with new elements and noted that roads and rivers can distort the rings. It was from this work that the concept of marginal cost (Grenzkosten) was put into practice, following the work of Alfred Marshall. Von Thünen’s model describes a socially efficient equilibrium; in fact, it is a fine example of Adam Smith’s concept of the ‘invisible hand’. This model is based on an isolated, homogeneous city without any territorial diversity. The industrial revolution and new developments have also had an impact on the applicability of the model (e.g., refrigerators allow farther transportation). Another problem is the lack of profit in this model, producers compete for the best locations, their strength measured through the concept of locational rent. However, after deducting production and transport costs, no profit remains, so the model assumes self-sufficiency of producers.

2. **More complex models – Weber**

In 1909, Alfred Weber published his work on the location theory of industrial activities [in German: *Über den Standort der Industrien*] in Tübingen (Weber, 1909). The aim of the model was to explain/predict the spatial distribution of industrial production – the explanation being that actors seek to minimise the costs of transport and labour. Weber assumed the existence of a single, free, and homogeneous market, a boom in industrial production, the development of transport infrastructure and the development of cities. The conditions of the model are:

1. one country, homogeneous topography, climate, technology, economic system;
2. the model considers one finished product at a time, supplied to the market, with infinite demand at a given price (which is uniform – more understandable in von Thünen, not so much here);
3. the raw materials and the market are in a known, fixed location;
4. labour is geographically fixed, but available everywhere in infinite quantities;
5. transport costs are directly proportional to the weight of the good and the distance.

It is typical of Weber’s time that transport costs were high (including wages) despite the development of infrastructure. Three criteria need to be considered to select the right location.

1. Transport cost – the ratio of the weight of the final product to the weight of the raw materials. (E.g., copper production is a weight loser – it would be expensive to transport ore to the market for processing, so the industry would be close to the raw material deposits. If the ratio is one, then it can be anywhere, if it is weight increasing, then it will be near the market.)

2. Labour (lower labour costs allow greater transport distances and could become a major factor in production). If the savings from cheaper labour outweigh the increased transport costs, it is worth deviating from the otherwise optimal location. This substitutability of wages and transport costs had a major impact on later models. In Weber’s time, there was no commuting.

3. Agglomeration (concentration of producers in a small area, which has advantages and can attract service industries). After calculating the agglomeration savings, isodapanes are drawn around the optimal location of each firm. If they overlap, agglomeration is appropriate. Weber distinguished between two cases: one in which a firm must be located far enough away to touch other markets and raw materials, and one in which the merger of plants at a point between their original sites is necessary to achieve economies of scale.

Note that Weber considered these aspects sequentially, not simultaneously, in his model (unrealistic). He transformed the difference in extraction of each raw material into a transport cost (the more difficult materials to extract are transported further). In Weber’s model, the problem is to find the point where the transport costs for the three factors (market, two types of raw materials) are minimum. Because of the importance of transport costs, Weber created a map to identify the location with the lowest cost, he called points with the same total transport cost isodapane (these are U-shaped black lines in Figure 2), the the line connecting points with the same transport cost for a single raw material is an isotim (white lines in Figure 2). This model does not consider loading costs, so it gravitates too much towards the locations between the raw materials and the market. The model also ignores spatial variations in demand, the real nature of labour (not unlimited, but mobile) and the fact that many market players produce several products from several raw materials for several markets, so the theory is not applicable.

3. Heckscher and Ohlin
Heckscher and Ohlin’s book of 1933 draws heavily on Ricardo’s work (Heckscher and Ohlin, 1933), starting from an analysis of economic relations between countries between which the factors of production (labour, capital) cannot flow, only the products produced are transported. In contrast to Ricardo’s model, which only takes labour into account as a factor of production (and thus can only generate comparative advantage based on technological differences), the Heckscher–Ohlin model also takes capital into account, and thus, unlike Ricardo, can ignore technological differences. Different endowments and factors of production are present in different proportions in different countries. Capital or labour shortages or surpluses cannot be equalised, because there are no factor flows between countries. As a consequence, scarce factors necessarily become more expensive, while abundant factors necessarily become cheaper. Since the costs of products are determined by the factors of production used, labour-intensive products will be cheaper to produce in labour-rich countries and capital-intensive products in capital-rich areas. Cheaper production also means greater competitiveness, so that exports of these products to countries where the factor of production is scarce will start or increase. Ohlin’s central idea is that if, for example, a labour-surplus country produces a high proportion of labour-intensive products and exports them, labour becomes a scarcer resource, leading to an increase in wages. For two countries and two products, the assumptions of the model are as follows.
1. Factors of production are not equally available in the two countries.

2. One of the products requires more capital and the other more labour.

3. Capital and labour do not flow between the two countries.

4. The transport of products between countries is costless.

5. Consumers in the two countries have the same needs.

6. The level of technology is the same in both countries.

The theory is not based on the total amount of capital and labour available, but on the amount per worker. A good example of this is 19th century England, which imported food from countries where land was abundant. Without it, England would have been forced to produce its own food, including areas with low fertility, thus significantly increasing the income of landowners (which led to the debate on grain tariffs). Trade between countries is therefore not simply an exchange of goods, but it also contributes to the equalisation of factor prices (wages, rents, interest), partly replacing labour and capital flows. It should be noted that even a completely free movement of factors would not fully compensate for differences, since, for example, the stock of buildings is not mobile, and there are also immutable geographical and climatic factors. In this model, labour is not a homogeneous factor of production, there are various categories of labour in terms of skills and occupation. The theory was refuted by Wassily Leontief, who pointed out that the United States is capital-rich, yet exports labour-intensive products (Leontief’s paradox). However, if labour is broken down into two factors (skilled/unskilled), the theory gives a more accurate prediction. The US imports products requiring unskilled labour and exports products requiring highly skilled labour.

4. Christaller – the theory of central places

Walter Christaller published his work on central place theory in 1933 (Christaller, 1933) – the first theory to attempt to treat cities as a system. After his entry into the Nazi Party, he was tasked with the economic reorganisation of the conquests in the East using his theory. The theory is based on the idea that cities are nothing more than central places that provide services to the surrounding region (Szabó et al., 2018). The assumptions of the theory are as follows:

1. flat topography, infinite space, evenly distributed population;

2. settlements are equidistant and located on a triangular grid;

3. resources are evenly distributed, there is only one type of transport, equally heavy in all directions;

4. distance decay mechanism: as distance increases, the number and strength of spatial interactions decreases;

5. perfect competition (no residual profit), everyone is profit-maximising (rational);

6. consumers have the same income level and shopping preferences, they try to minimize travel time (choose the nearest provider).

It follows from the above that the size of the regions providing a product or service is the same, transport costs are proportional to distance (Andrejszki et al., 2015). The theory is further based on two concepts: critical population/income (threshold – the level required to trade a product or service) and range (the distance consumers are willing to travel, above a certain distance the inconvenience outweighs the need for the good). Under these conditions, centres of varied sizes emerge, supplying assorted products, forming a hierarchical system that can be generalised. The larger the size, the fewer such centres there will be (many small villages, few large towns). With size, the distance between centres decreases (large towns are far apart, villages are close). With settlement size, the range of functions served and the number of higher-order services increases. The higher the level of a product/service (more durable/valuable/varied), the further consumers are willing to travel to obtain it. At the bottom of the hierarchy are e.g., bakeries, post offices, grocery stores, which are small in size, and at the top are e.g., jewellers, large shopping centres. The resulting system is arranged in a triangular space (as transport is equally difficult in all directions, the markets have a circular attraction, which is best served by a hexagonal system without overlaps). Christaller distinguished seven levels of settlement. Christaller gave different K-values for the spatial location of the centres.
According to the market rule, each higher-order centre occupies 1/3 of the market area of the neighbouring lower-order centre, so \( K = 3 \times (1+6^{*}1/3) \). The centres are then located at the vertices of the hexagons. Christaller noted that this system is not ideal from a transport point of view, as there is no direct connection between the larger centres; and therefore, proposed the transport rule \( (K = 4) \). In this case, the higher-order centre covers half the market area of the neighbouring lower-order centres \( (1+6^{*}1/2) \), with the centres located on the sides of the hexagons. This arrangement minimizes the length of the roads. From a political and administrative point of view, this arrangement is not ideal, as the overlapping of scopes is not possible in this case. This problem is solved by the administrative rule \( (K = 7) \), according to which the higher-order centre covers the whole of the neighbouring lower-order centres \( (1+6^{*}1) \). The model explains well the phenomenon of urbanisation, the hierarchy of urban centres and the spatial distribution of trade and services. In reality, large flat areas are rare, the location of industries can be influenced by political factors, perfect competition is unrealistic, consumers’ shopping habits are diverse (not necessarily optimised for travel time minima), consumers and resources are never evenly distributed in space.

5. Lösch’s theories

The world wars and the Great Depression (1929–1933) highlighted the importance of territorial market analyses and the need to further develop theories of establishment. These new directions were summarised by August Lösch in his 1940 work *Die räumliche Ordnung der Wirtschaft* (The Spatial Order of the Economy), which drew heavily on Christaller’s work of 1933 (Lund, 1943). In contrast to Christaller, Lösch started from the lowest, rather than the highest, order of places (subsistence farms), which he divided into a grid of triangles and hexagons. From the lowest level of economic activity, Lösch developed a number of systems based on central places, including Christaller’s 3 systems. But Lösch’s theory included specialised places and showed how some regions become richer than others. Lösch deviated from Christaller’s model in the following ways:

1. the amount of area supplied by higher levels to lower levels is not constant;
2. higher level settlements do not necessarily have the same functions as lower-level settlements;
3. specialised centres are more likely to develop near the centre of an area than further away from it.

The model gives a general equilibrium of locations and price levels described by systems of equations.

6. Harold Hotelling – linear markets

Harold Hotelling (1990) established the theory of linear markets, with the following premises.

1. A linear city is assumed.
2. Consumers are evenly distributed.
3. There are two firms in the area selling the same product at the same price, the only difference between them is their location. Changing the location is costless.
4. Customers are also homogeneous, preferring neither firm.

In this situation, profit can be maximised by increasing the number of customers a firm serves. Thus, the firms move towards each other, until both are at the halfway line. So, in the case of a vendor, it is advisable to place the ice cream van in the middle of the beach (minimum walking distance for consumers). For two vendors, the social equilibrium for both vendors is halfway between the bisector line and the end of the beach (dividing the space into 25% sections). However, this is not a Nash equilibrium: if one vendor moves closer to the other, it may attract a portion of the other’s customers, so they both end up in the middle line of the beach. At that point, neither can improve their position by moving further (Nash equilibrium, but not socially optimal). Thus, in the presence of imperfect competition, the positioning of industries can be explained by the search to reach consumers and reduce competition. This is also reflected in practice (e.g., when a Starbucks is next to a Costa), the choice of close locations protects firms from aggressive competition.
7. Myrdal – circular causality

Myrdal’s theory is based on the fact that people prefer to go where there are lots of opportunities. Companies prefer to move to a place where there are many potential employees. This is based on their utility and profit maximising behaviour. Myrdal with his model started to explain regional differences (Myrdal, 1957). He examined the so called “Manufacturing belt” – where there is a large market but with mutually reinforcing effects (Harris, 1954) Consumer and business decisions can reinforce each other. Historical coincidences can also explain the location of industries. Based on his theory, it can be understood why competing companies share location.

8. Paul Krugman – the new economic geography

In 1991 in two papers, Paul Krugman examined the question of when and under what conditions industrial activity concentrates in some regions, while other regions develop more slowly and thus lag behind (Krugman, 1991). In short, his theory states that economies of scale favour larger firms because they can produce more cheaply than smaller firms entering the market. Thus, early entrants have an advantage, and production is easily concentrated, creating monopolistic or oligopolistic conditions. It is the governments’ task to stop these processes.

In the basic model of the new economic geography, there are two regions of various levels of development – North and South, in Krugman’s terms. In both regions, there are two sectors of production activity: in the agricultural sector, an unchanging number of workers working exclusively in that sector produce agricultural goods under conditions of perfect competition (they cannot move); in the industrial sector, an unchanging number of workers, but free to move between sectors, produce industrial goods in a large number of firms using technology with increasing returns. There is one agricultural product the price of which is the same everywhere, the cost of transport is zero, one region exports exclusively, the other one imports exclusively.

The movement of industrial workers between regions depends on the real wage level. Monopolistic conditions of competition prevail in the market for industrial goods. The transport of industrial goods raises costs, which makes imported goods from other regions more expensive than their own goods. (Krugman neglects intra-regional transport, treating regions as a one-point economy in themselves.) There is a demand for all goods produced, but the extent of this demand varies. There is no difference in consumer preferences. So, an asymmetric structure, a centre-periphery structure emerges. However, competition between the growing number of entrepreneurs in the richer region is increasing, which makes their counterparts already there look for a new location, as they will be more interested in starting their production activities in a less developed region, where competition is less intense. There, in turn, real wages are lower, demand is lower, and incomes will be lower. A centre-periphery structure will therefore emerge if the entrepreneur seeking to locate in the area perceives the benefits of agglomeration to be greater than those of less competition.

Apart from a moment from the benefits of smaller competition, the process of divergence will be triggered if, for whatever reason, more workers are located in a given region (Csete and Szabó, 2014). Demand is then higher, attracting producers, who have higher incomes, higher real wages, and more industrial workers in the region in question. But this brings us to some problematic parts of Krugman’s concept. First, there are several links between the immigration of industrial workers and the level of regional real wages.

1. The larger the regional market is, the more industrial products are produced there, which has a positive effect on the price level of industrial goods and thus on real wages.
2. The smaller the regional market, the less the competition. In this case, the entrepreneur is able to charge higher prices because of the relative or absolute excess demand for his products, which in turn reduces the real wage.
3. In contrast, a larger regional market leads to an oversupply – relative or absolute – which lowers prices and therefore raises real wages.

The final result depends on which of the three real wage effects is (are) stronger. Bearing the above cases in mind, it can be seen that the agglomeration process will be stronger if the following conditions are met.

1. Industrial goods are difficult to substitute, because it is difficult to replace the products that consumers demand and import with products produced locally, i.e., it is not worth relying on local production.
2. The weight of manufactured goods in the determination of the price index is high, because in this case, the price level falls, and real wages rise.

The other highly controversial part of Krugman’s model is the starting point: if there is even a small difference in development between the two regions, then the gap between the two regions will widen if there are low transport costs, difficult substitutes and/or different numbers of industrial goods and workers, as described above. But Krugman does not explain why these differences occur. So, if a region happens (!) to be in a more disadvantaged position, it has no chance of catching up if the above conditions are met. This brings us to the economic policy implications of the theory under discussion.
In Krugman’s model, the spatial concentration of production reduces transport costs for all participants, and these costs are constantly falling as a result of technological progress. This theorem could be generalised and applied even for road traffic accident analysis (Sipos et al., 2021). The trends in economic location point towards further concentration and a worldwide increase in economic disparities, widening the gap between the centre and the periphery. Krugman has also found a solution to another trend, namely the spectacularly rapid development of some emerging countries. Weber’s model suggests that potential savings from wage differentials may justify a shift away from optimal location, determined by transport costs. Krugman points out that this may become increasingly common as transport costs fall worldwide.

It would obviously be superficial to say that he has not really added anything new to any of these trends, but that his contribution is a synthesis of existing results. However, one would not be too far from the truth; indeed, the methodology used, and the partial results obtained were already known before Krugman. Like his predecessors, he based his theory on micro-actors making rational decisions, more consistently with modern microeconomic approaches, thanks also to the advances in economics that have taken place in the meantime. An analogy can also be drawn with the use of external factors. In the earliest models (von Thünen, Weber), the sources of raw materials were – rightly, of course – considered to be economically exogenous, and even the supply of factors was determined in this way. The dynamics of the system were determined by the transport of products and raw materials. Krugman, on the other hand, assumes the complete mobility of industrial workers, but, as we have seen, he linked the beginning of the dynamics to the regional distribution of labour; the regional structure of industry is transformed when the distribution of industrial workers is not even or when there is a disturbance in this respect.

At this point, one could say that von Thünen’s and Weber’s models rely overwhelmingly on the technological context, with the demand side being a passive condition, whereas Krugman takes this side into account through the evolution of real wages, as an active factor in the process. This, however, is already to be found in the works of Lösch and Christaller. Indeed, they managed to derive the settlement structure of the national economy by referring to it, just as Krugman derived the formation of agglomerations and the centre-periphery structure.

In all the earlier models, as in the basic model of the new economic geography, there is a sense of technological development or, to move on to a comparison with growth theory, the key role of human capital for the division of labour between regions (Szabó et al., 2017). If new knowledge is acquired in the national economy under study, not only is the efficiency of production improved, but this will certainly also have an impact on transport costs. In turn, this may change the previously established structure of sites or settlements. In Krugman’s model, too, technical progress implies important but by no means obvious changes. On the one hand, product diversification will increase, making substitutability more difficult and, as mentioned above, reinforcing the divergent process. On the other hand, if technical progress is concentrated in the less developed region, the competitive position of the South will improve, thus increasing the income generated there and giving this region a chance to catch up.

9. Spatial impossibility theorem

The spatial impossibility theorem assumes a finite number of producers and consumers at a finite number of locations. If space is homogeneous, transport is costly and preferences are locally non-satiated, then there is no competitive equilibrium with transport costs. In other words, the price system must perform two functions: it must support trade between regions, and it must prevent firms and households from relocating to other regions. The spatial impossibility theorem states that both objectives are impossible to achieve in a homogeneous space, as pro-competitive prices are bad for territorial stability. E.g., if a product is produced in region A and exported to region B, producers in region A will want to move to region B for higher profits, while buyers in region B will want to move to region A for lower prices. The basic shortcomings of the spatial impossibility theorem for competition theories, and its ultimate conclusion, is that in a homogeneous space with transport costs, the only possible competitive equilibrium is the so-called back garden capitalism, where each location is autarkic. In other words, simple assumptions and a simple competitive market model are not enough, to understand the spatial location of economic activities. We need to accept at least one of the following three assumptions:

1. local externalities (Marshall);
2. heterogeneous space (von Thünen) + external trade (Heckscher–Ohlin);
3. imperfect competition (Krugman).

The model of Dixit and Stiglitz (1977) explaining the equilibrium of the economy under conditions of monopolistic competition is a landmark in the explanation of space, as it allows the development of a centre-periphery model. Thomas Piketty’s Capital in the Twenty-First Century (2014) argues that under capitalism, the rich get richer and the poor get poorer, undermining the entire system in the long run. The reason is that the return on capital – which people earn as annuities on the property, shares, and other assets they own – grows faster than the economy as a whole expands, so wealth is always concentrated.
Acknowledgement
Author is deeply acknowledge the support and advice of Ms Eva KOZAK.

References


